系 统 科 学 与 数 学
J. Sys. Sci. & Math. Scis.
28(12) (2008, 12), 1525–1534

带 NCP 函数的信赖域滤子方法*

苏 珂

(河北大学数学与计算机学院，保定 071002; 同济大学数学系，上海 200092)

摘要 滤子方法最初是由 Fletcher 和 Leyffer 在 2002 年提出的. 这种方法的原理是：在一个试探步，如果相应的目标函数值或约束违反度函数值下降，那么该试探步就会被接受. 利用 Fischer-Burmeister NCP 函数来修正滤子中的约束违反度函数，同时证明了这个新的滤子方法具有全局收敛性.

关键词 滤子，信赖域，非线性互补问题，收敛.

MR(2000) 主题分类号 90C30

1 引 言

考虑如下非线性规划问题

$$\text{(P)} \qquad \begin{array}{ll} \min & f(x) \\ \text{s.t.} & g_i(x) \le 0, \quad i \in I = \{1, 2, \cdots, m\}, \end{array}$$

其中 $x \in R^n, f: R^n \to R$, $g(x) = (g_1(x), g_2(x), \cdots, g_m(x))^{\mathrm{T}}: R^n \to R^m$ 均为二次连续可微实值函数.

很多对于问题 (P) 的求解方法都依赖于牛顿方法且为迭代算法. 即给定问题 (P) 的最优解 x^* 的一个估计 x^k , 求解问题 (P) 的一个线性二次逼近来求得新的比 x^k 更优的估计 x^{k+1}. 而在实际运算过程中，如果初始点在最优解的附近，那么这个过程保证了很好的收敛性，但是如果初始点在远离最优解的地方，那么这个过程产生的 $\{x^k\}$ 可能是不收敛的. 如何保证从远离最优解的地方开始迭代所得到的序列是收敛的呢？传统意义上讲，这个问题已经得到了回答，那就是使用罚函数或者价值函数. 这类函数是目标函数和约束违反度函数的一种线性组合. 其中约束违反度函数的形式为 $h(x) = \|g(x)^+\|$, 这里 $g_i^+ = \max\{g_i, 0\}$, $\|\cdot\|$ 表示欧几里德范数. 例如 l_1 精确罚函数 $\pi(x, \rho) = f(x) + \rho h(x)$, 其中 $\rho > 0$ 为罚因子，只要 ρ 充分大，我们就可以使用这个罚函数在每步迭代充分下降的基础上保证我们迭代算法的运行.

不幸的是，一个合适的罚因子是依赖于问题 (P) 的解的. 即常存在一个临界值，若罚因子小于此临界值，则罚函数在问题 (P) 的解处没有局部最小值. 而此临界值先前是不知道的. 在计算中，如果罚因子选择的太小，则可能得到问题 (P) 的不可行点或者造成惩罚项的无限增大. 另一方面，如果罚因子选择的太大，则会降低目标函数的作用和影响，导致收敛

* 国家自然科学基金 (10771162) 资助课题.
收稿日期：2006-05-15, 收到修改稿日期：2007-12-21.

过慢. 于是为了避免罚因子的选取, Fletcher 和 Leyffer 提出了滤子方法, 这种方法是通过比较约束违反度函数值和目标函数值来决定试探步是否被接受. 一个新的试探点被滤子接受, 当且仅当它的约束违反度函数值下降或者目标函数值和所有滤子中相应的函数值相比有充分的下降. 近年来, 由于滤子方法良好的数值结果 [1], 对它的讨论也越来越多.

在滤子概念提出后, 这个技巧应用到了诸如 SQP[2] 、线搜索 [3] 、信赖域 [4] 等很多方法. 之后 Ulbrich[5] 和 Fletcher 等 [6] 又分别证明了 SQP 滤子方法的局部和全局收敛性, Fletcher, Leyffer 和 Toint[7] 又提出了 SLP 滤子算法的全局收敛性.

后来滤子方法得到了不断改善和发展, 聂普焱曾在文 [8] 中加入参数, 将滤子方法中的滤子对进行松弛从而得到了一类新的算法; 又在文 [9] 中将滤子方法和试探步分解技术相结合. Peng 在文 [10] 中将滤子方法与无梯度方法相结合. 在这些方法中都只是利用了目标函数和约束函数的性质, 而在本文中, 我们提出另外一种算法, 仍然属于滤子方法, 只是和以前工作不同的是, 利用 NCP 函数来构造滤子, 并与信赖域技术相结合, 同时得到了该算法的全局收敛性.

2 Fischer-Burmeister NCP 函数和滤子技巧

非线性规划问题 (P) 的 Lagrangian 函数为

$$L(x,\lambda)=f(x)+\lambda^{\mathrm{T}}g(x)=f(x)+\sum_{i=1}^{m}\lambda_i g_i(x), \tag{1}$$

其中 $\lambda=(\lambda_1,\lambda_2,\cdots,\lambda_m)^{\mathrm{T}}\in R^m$ 是乘子向量. 在不引起混淆的情况下, 用 (x,λ) 来表示向量 $(x^{\mathrm{T}},\lambda^{\mathrm{T}})^{\mathrm{T}}$.

我们称满足如下条件的点 $(x,\lambda)\in R^{n+m}$ 为问题 (P) 的 Karush-Kuhn-Tucker(KKT) 点

$$\begin{gathered}\nabla_x L(x,\lambda)=0,\\ g_i(x)\le 0,\quad \lambda_i\ge 0,\quad \lambda_i g_i(x)=0,\quad i\in I.\end{gathered} \tag{2}$$

为了求解 (2) 以得到问题 (P) 的 KKT 点, 我们引入 NCP 函数.

定义 2.1 函数 $\phi: R^2\to R$, 如果它具有

$$\phi(a,b)=0 \Longleftrightarrow a\ge 0,\quad b\ge 0,\quad ab=0,$$

则称 ϕ 为一个 NCP 函数.

较为常用的 NCP 函数之一为 Fischer-Burmeister NCP 函数, 简记为 F-B NCP 函数, 它具有如下形式

$$\phi(a,b)=\sqrt{a^2+b^2}-a-b.$$

利用 F-B NCP 函数, KKT 条件 (2) 变为

$$\begin{gathered}\nabla_x L(x,\lambda)=0,\\ \Phi(x,\lambda)=0,\quad i\in I.\end{gathered} \tag{3}$$

其中 $\Phi(x,\lambda)=(\phi(-g_1(x),\lambda_1),\phi(-g_2(x),\lambda_2),\cdots,\phi(-g_m(x),\lambda_m))^{\mathrm{T}}$.

滤子概念由 Fletcher 和 Leyffer 于 2002 年提出 [1], 它的原理是要考察目标函数值和约束违反度函数值. 他们使用的约束违反度函数为

$$p(x) = \|g(x)^+\|_\infty,$$

其中 $g_i(x)^+ = \max\{0, g_i(x)\}$ $(i \in I)$. 易见若约束违反度函数 $p(x) = 0$ 则意味着 x 是问题 (P) 的可行点.

定义 2.2 称点 x^k 支配点 x^l, 当且仅当 $p^k \le p^l$ 且 $f^k \le f^l$, 其中 $p^k = p(x^k)$.

定义 2.3 滤子就是具有形式 (p, f) 的列表，记为 $\mathcal{F}$, 使得

$$\text{或者 } p(x^i) \le p(x^j) \text{ 或者 } f(x^i) \le f(x^j)$$

对所有 $i \ne j$ 成立.

为了提供一种有效的算法工具，我们需要定义试探点不被接受的准则. 事实上，如果迭代点 x^{k+1} 的对应函数对的值 (p^{k+1}, f^{k+1}) 非常接近 x^k 对应的值或接近已经存在于滤子中的值，它就会被滤掉. 这意味着在 (p, f) 平面上，我们对不被滤子接受的点所在的区域建立了一个小的边界. 因此，一个试探点 x 被滤子接受，当且仅当

$$p(x) \le \beta p(x^l), \tag{4}$$

或

$$f(x) \le f(x^l) - \gamma p(x), \tag{5}$$

$\forall (p(x^l), f(x^l)) \in \mathcal{F}$ 成立. 其中 $0 < \gamma < \beta < 1$ 为给定的正常数.

需要说明的是，滤子集合是不断变化的，我们用 $\mathcal{F}_k$ 来表示第 k 次迭代时的滤子集合. 那么随着算法的进行，我们需要将点对 (p, f) 加入滤子中. 如果 $x^k + d^k$ 被当前滤子 $\mathcal{F}_k$ 接受，那么记 $x^{k+1} = x^k + d^k$, 并且

$$D_{k+1} = \{(p^j, f^j) | p^j \ge p^k \text{ 且 } f^j \ge f^k, \ \ \forall (p^j, f^j) \in \mathcal{F}_k\}.$$

滤子集合按下述方法更新

$$(\mathcal{F}_{k+1}) \ \mathcal{F}_{k+1} = \mathcal{F}_k \bigcup \{(p^{k+1}, f^{k+1})\} \setminus D_{k+1}. \tag{6}$$

我们也称这步操作为 “将 $x^k + d^k$ 加入滤子”, 虽然, 严格来讲, 是把这一点对应的 (p, f) - 对加入滤子集合.

3 算 法

为了充分利用 NCP 函数与 KKT 条件间的关系和 NCP 函数几乎处处可微的性质，我们改造滤子中的度量，将约束违反度函数 $p(x)$ 修正为

$$p(x, \mu) = \|\varPhi(x, \mu)\|_\infty = \max_{1 \le i \le m} |\phi(-g_i(x), \mu_i)|,$$

其中 $\phi(-g_i(x), \mu_i) = \sqrt{g_i^2(x) + (\mu_i)^2} + g_i(x) - \mu_i$. 点 x^k 处对应的函数 p 和 f 的值简记为 (p^k, f^k).

借鉴文献 [1] 中的思想，为避免 p 值过大的点进入滤子，我们用一个上界作为接受迭代点的必要条件

$$p(x,\mu)\le u.$$

其中 u 为一大的正数．于是初始化滤子集合为 $\mathcal{F}_0=\{(u,-\infty)\}$.

对于一个给定的迭代点 $x^k\in R^n$, 即问题 (P) 的局部最优解 x^* 的当前估计，我们解如下问题 $(QP(x^k))$ 得 d^k(同时可以得到问题 $(QP(x^k))$ 对应于 d^k 的乘子 λ^k).

$$\begin{aligned}(QP(x^k))\qquad\qquad & \min\quad q^k(d)=d^{\mathrm T}\nabla f(x^k)+\frac{1}{2}d^{\mathrm T}H^kd\\ & \text{s.t.}\quad g_i(x^k)+d^{\mathrm T}\nabla g_i(x^k)\le 0,\quad i\in I.\end{aligned}\tag{7}$$

其中 H^k 是拉格朗日函数的近似 Hessian 阵.

不幸的是，由于牛顿迭代的局部收敛性，迭代 d^k 可能很不好．克服此困难的一个方法是定义一个价值函数，随着 d^k 转好该函数值下降，信赖域或线搜索方法均可以用来最小化价值函数，同时在合理的假设下保证整体收敛性．但是为了避免使用罚函数，我们来考虑不带惩罚函数的信赖域方法．具体做法是在范数意义下限制迭代步 d^k 使得 x^k+d^k 保持在以 x^k 为中心的信赖域范围内．即用如下定义的问题 $QP(x^k,\triangle^k)$ 来代替问题 $QP(x^k)$

$$\begin{aligned}(QP(x^k,\triangle^k))\qquad\qquad & \min\quad q^k(d)=d^{\mathrm T}\nabla f(x^k)+\frac{1}{2}d^{\mathrm T}H^kd\\ & \text{s.t.}\quad g_i(x^k)+d^{\mathrm T}\nabla g_i(x^k)\le 0,\quad i\in I,\\ & \qquad\ \|d\|\le\Delta^k.\end{aligned}\tag{8}$$

其中 $\|\cdot\|$ 表示欧几里得范数．需要指出的是，若 x^k+d^k 不能够被接受，但是却很接近最优解，即出现了所谓的 Maratos 效应．为了克服此现象，我们在 x^k+d^k 不被滤子接受的时候去计算二阶校正步 (SOC), 即解如下子问题

$$\begin{aligned}(QP_s(x^k,\triangle^k))\qquad\qquad & \min\quad (d^k+d)^{\mathrm T}\nabla f(x^k)+\frac{1}{2}(d^k+d)^{\mathrm T}H^k(d^k+d)\\ & \text{s.t.}\quad g_i(x^k+d)+d^{\mathrm T}\nabla g_i(x^k)\le 0,\quad i\in I,\\ & \qquad\ \|d^k+d\|\le\Delta^k.\end{aligned}\tag{9}$$

其中 d^k 是问题 $(QP(x^k,\triangle^k))$ 的解．得解 $\overline{d}^k$, 于是下一步迭代为 $x^{k+1}=x^k+d^k+\overline{d}^k$.

令 $\Delta f^k(d)=f(x^k)-f(x^k+d^k)$ 表示函数 $f(x^k)$ 的真实下降量，

$$\Delta q^k(d)=q^k(0)-q^k(d)=-d^{\mathrm T}\nabla f(x^k)-\frac{1}{2}d^{\mathrm T}H^kd$$

表示 $f(x^k)$ 的预测下降量． $f(x^k)$ 的充分下降条件表示为 $\Delta f^k(d)\ge\tau\Delta q^k(d)$, 其中 τ 是某常数.

算法 A

S0 选取初值点 x^0, 给定初始信赖域半径 $\Delta^0>0$, 初始对称阵 H^0, 常数 $\mu^0\ge 0$, $\tau\in(0,\frac{1}{2})$, $0<\gamma<\beta<1$, $p^0=p(x^0,\mu^0)$, $k=0$. 设初始滤子集合为 $\mathcal{F}_k=\{(u,-\infty)\}$, 其中 $u=M\max\{1,p^0\}$, $M>0$;

S1 求解 $QP(x^k,\Delta^k)$, 若问题不可行，则进入可行性恢复阶段，找到点 x^{k+1} 使得 $QP(x^{k+1},\Delta^{k+1})$ 是可行的． $k=k+1$, 并得解 d^k;

S2 若 $d^k=0$, 停止，得到 KKT 点 x^k. 否则计算拉格朗日乘子估计 λ^k. 记 $x_+^k=x^k+d^k,\Delta f^k=f(x^k)-f(x^k+d^k),\Delta q^k=-\nabla f(x^k)^{\mathrm T}d^k-\frac{1}{2}(d^k)^{\mathrm T}H^kd^k$;

S3 若 (x_+^k,λ^k) 被 $\mathcal{F}_k\bigcup(x^k,\mu^k)$ 接受，则转 S4, 否则转 S5;

S4 若 $\Delta f^k<\tau\Delta q^k$ 且 $\Delta q^k>0$ 成立，转 S6, 否则转 S7;

S5 计算二阶校正步

S5.1 求解 QP_s 得 $\overline{d}^k,\overline{\lambda}^k$, 记 $\overline{x}^k=x^k+d^k+\overline{d}^k,\Delta\overline{f}^k=f(x^k)-f(x^k+d^k+\overline{d}^k),\Delta\overline{q}^k=q^k(0)-q^k(d^k+\overline{d}^k)$;

S5.2 若 $(\overline{x}^k,\overline{\lambda}^k)$ 被 $\mathcal{F}_k\bigcup(x^k,\mu^k)$ 接受，转 S5.3, 否则转 S6;

S5.3 若 $\Delta\overline{f}^k<\tau\Delta\overline{q}^k$ 且 $\Delta\overline{q}^k>0$ 成立，则转 S6, 否则转 S5.4;

S5.4 $x^{k+1}=x^k+d^k+\overline{d}^k,\mu^{k+1}=\overline{\lambda}^k$. 若 $\Delta\overline{q}^k\le 0$, 则将 $(\overline{x}^k,\overline{\lambda}^k)$ 加入滤子，转 S8;

S6 $x^{k+1}=x^k,\mu^{k+1}=\mu^k,\Delta^{k+1}=\frac{\Delta^k}{2},k=k+1$, 转 S1;

S7 $x^{k+1}=x^k+d^k,\mu^{k+1}=\lambda^k,\Delta^{k+1}=\frac{\Delta^k}{2}$, 若 $\Delta q^k\le 0$, 则将 $(\overline{x}_+^k,\lambda^k)$ 加入滤子，转 S8;

S8 更新矩阵 H^k 为 H^{k+1}, $k=k+1$, 转 S1.

在上述算法中，可行性恢复阶段的目的是使子问题 $QP(x^k,\Delta^k)$ 可行，其方法可参见 Fletcher[1] 中的 SLP 恢复阶段方法．所以经过可行性恢复阶段总可以找到点 x^k 使得 $QP(x^k,\Delta^k)$ 可解．矩阵 H^k 的更新可以采用 BFGS 校正公式．

从算法中我们可以看出，并不是所有被滤子接受的点都被加入滤子集合，只有当 $\Delta q^k\le 0$ 时，迭代点才被加入滤子，此时 f 的值可能增大；而当 f 的值下降良好时，并不将迭代加入滤子．

引理 3.1 设序列 p^k 和 f^k 满足 $p^k\ge 0$ 且 f^k 单调递减下有界，令常数 β,γ 满足 $0<\gamma<\beta<1$. 若对于任意的 k, 或有 $p^{k+1}\le\beta p^k$, 或有 $f^k-f^{k+1}\ge\gamma p^{k+1}$ 成立，则 $\lim\limits_{k\to\infty}p^k=0$.

证 文献 [2] 中引理 1.

引理 3.2 设无穷序列 (p^k,f^k) 进入滤子，其中 $p^k>0$, $\{f^k\}$ 下有界，则有 $\lim\limits_{k\to\infty}p^k=0$.

证 文献 [2] 中推论.

需要指出的是，若有无穷多点进入滤子，则必有 $p^k>0$ 对充分大的 k 成立．事实上，若 $p^k=0$, 而 x^k 不是 KKT 点的情况下，必有 $p^k>0$, 由算法可知这样的点是不能加入滤子的，矛盾．故有 $\min\limits_{(p^l,f^l)\in\mathcal{F}_k}p^l>0$.

4 收敛性分析

在下面的分析中，需要如下一些假设

A1 目标函数 f 和约束函数 $g_i(x),i\in I$ 都是二次连续可微的.

A2 迭代 $(x^k,\mu^k),(x^k+d^k,\lambda^k),(x^k+d^k+\overline{d}^k,\overline{\lambda}^k)$ 均位于 R^{n+m} 的有界紧凸集 S 内.

A3 存在正常数 a,b 使得 Hessian 阵 H^k 满足 $a\|d\|^2\le d^{\mathrm T}H^kd\le b\|d\|^2$, $\forall k$ 和 $d\in R^n$ 成立.

A4 子问题 $QP(x^k,\Delta^k)$ 的 KKT 乘子 λ^k 一致有界，$QP_s(x^k,\Delta^k)$ 的 KKT 乘子 $\overline{\lambda}^k$ 一致有界.

由假设 A1, A2 可知，对任意的 k, 有

$$f^{\min}\le f(x^k)\ \text{且}\ 0\le p^k\le p^{\max}, \tag{10}$$

其中 $f^{\min}>0, p^{\max}>0$ 为两常数. 故在 (p,f) 平面被滤子接受的 (p,f) 位于矩形 $[0,p^{\max}]\times[f^{\min},\infty]$ 内. 同时由 A1–A4 知，存在 $M>0$, 使得

$$\|H^k\|\le M,\quad \|\nabla^2 f(x)\|\le M,\quad \|\nabla^2 g_i(x)\|\le M,\quad i\in I,\quad \|\lambda^k\|\le M,\quad \|\overline{\lambda}^k\|\le M.$$

由上述说明，恢复阶段总是成功的，分析算法可知，算法 A 只存在如下 3 个使迭代不成功 (即 $x^{k+1}=x^k$) 的循环.

1) 循环 1 S1-S2-S3-S5.1-S5.2-S6-S1;

2) 循环 2 S1-S2-S3-S5.1-S5.2-S5.3-S6-S1;

3) 循环 3 S1-S2-S3-S4-S6-S1.

为了说明算法 A 是有效的，迭代是成功的，我们有必要说明上述 3 个循环是有限终止的. 这体现在下面 3 个引理中.

引理 4.1 算法 A 中循环 1 有限终止，即当 k 充分大时，$(x^k+d^k+\overline{d}^k,\overline{\lambda}^k)$ 必被 $\mathcal{F}_k\cup(x^k,\mu^k)$ 接受.

证 为书写简便，记 $\widehat{d}^k=d^k+\overline{d}^k,\overline{x}^k=x^k+\widehat{d}^k,\overline{p}^k=p(\overline{x}^k,\overline{\lambda}^k),\overline{f}^k=f(\overline{x}^k)$, 分两部分证明结论.

1) k 充分大时，需证明 $(\overline{x}^k,\overline{\lambda}^k)$ 可被 (x^k,μ^k) 接受. 即满足

$$\overline{p}^k\le\beta p^k, \tag{11}$$

或

$$f^k-\overline{f}^k\ge\gamma\overline{p}^k. \tag{12}$$

假设循环非有限终止，由算法知 $\exists\{k\}\subset K(K$ 为无限集$)$, 使得 $\Delta^k\to0(k\in K,k\to\infty)$, 而由 $\|\widehat{d}^k\|\le\Delta^k$, 知 $\widehat{d}^k\to0(k\in K,k\to\infty)$. 考虑如下两种情形.

情形 A $p^k=p(x^k,\mu^k)=0$.

此时有 $\|\Phi(x^k,\mu^k)\|=0$, 于是 $\forall i$ 有 $\phi(-g_i(x^k),\mu_i^k)=0$. 由 ϕ 定义知

$$g_i(x^k)\le0,\quad \mu_i^k\ge0,\quad \mu_i^kg_i(x^k)=0,\quad i\in I.$$

下证 (12) 成立.

$$\begin{aligned}\overline{p}^k&=p(x^k+\widehat{d}^k,\overline{\lambda}^k)\\&=\max_{i\in I}\left|\sqrt{g_i^2(x^k+\widehat{d}^k)+(\overline{\lambda}_i^k)^2}+g_i(x^k+\widehat{d}^k)-\overline{\lambda}_i^k\right|\\&=\max_{i\in I}\Big|\sqrt{[g_i(x^k)+\nabla g_i(x^k)^{\mathrm T}\widehat{d}^k+O(\|\widehat{d}^k\|^2)]^2+(\overline{\lambda}_i^k)^2}\\&\quad+g_i(x^k)+\nabla g_i(x^k)^{\mathrm T}\widehat{d}^k+O(\|\widehat{d}^k\|^2)-\overline{\lambda}_i^k\Big|.\end{aligned} \tag{13}$$

由 $\overline{\lambda}_i^k$ 定义知

a) 当 $g_i(x^k)=0$ 时，或者有 $\overline{\lambda}_i^k>0, \nabla g_i(x^k)^{\mathrm{T}}\widehat{d}^k=0$, 或者有 $\overline{\lambda}_i^k=0, \nabla g_i(x^k)^{\mathrm{T}}\widehat{d}^k\le 0$. 对前者结合 (13) 式有 $\phi(-g_i(\overline{x}^k),\overline{\lambda}_i^k)=O(\|\widehat{d}^k\|^2)$; 后者直接有 $\phi(-g_i(\overline{x}^k),\overline{\lambda}_i^k)=0$.

b) 当 $g_i(x^k)<0$ 时,对 $\overline{\lambda}_i^k>0$, 有 $g_i(x^k)+\nabla g_i(x^k)^{\mathrm{T}}\widehat{d}^k=0$, 于是由 (13) 有 $\phi(-g_i(\overline{x}^k),\overline{\lambda}_i^k)=O(\|\widehat{d}^k\|^2)$；对 $\overline{\lambda}_i^k=0$, 有 $g_i(x^k)+\nabla g_i(x^k)^{\mathrm{T}}\widehat{d}^k\le 0$, 于是直接有 $\phi(-g_i(\overline{x}^k),\overline{\lambda}_i^k)=0$.

结合 a) 和 b) 可得此时有 $\overline{p}^k=O(\|\widehat{d}^k\|^2)$, 又由 $\widehat{d}^k$ 的定义知，必存在 $\gamma>0$ 使得

$$f^k-\overline{f}^k=f(x^k)-f(x^k+\widehat{d}^k)=-\nabla f(x^k)^{\mathrm{T}}\widehat{d}^k+O(\|\widehat{d}^k\|^2)>\gamma\overline{p}^k. \tag{14}$$

情形 B $p^k=p(x^k,\mu^k)\neq 0$.

下证此时 (11) 成立.

$$p^k=p(x^k,\mu^k)=\max_{i\in I}\left|\sqrt{g_i^2(x^k)+(\mu_i^k)^2}+g_i(x^k)-\mu_i^k\right|. \tag{15}$$

若 $p^k\neq 0$ 则必 $\exists i$ 有 $g_i(x^k)>0,\ \mu_i^k\ge 0$ 或 $g_i(x^k)<0,\ \mu_i^k>0$. 于是，对前者有

$$2g_i(x^k)\ge\sqrt{g_i^2(x^k)+(\mu_i^k)^2}+g_i(x^k)-\mu_i^k\ge g_i(x^k). \tag{16}$$

对后者有

$$-g_i(x^k)\ge|\sqrt{g_i^2(x^k)+(\mu_i^k)^2}+g_i(x^k)-\mu_i^k|\ge 0. \tag{17}$$

若 $\lambda_i^k=0$, 由 (13) 知

$$\phi(-g_i(\overline{x}^k),\overline{\lambda}_i^k)=0. \tag{18}$$

或若 $\lambda_i^k>0$, 有

$$\phi(-g_i(\overline{x}^k),\overline{\lambda}_i^k)=O(\|\widehat{d}^k\|^2). \tag{19}$$

故由 (13),(16)–(19) 知，当 $k\in K,\ k\to\infty$ 时，必 $\exists\beta$, 使 $\overline{p}^k\le\beta p^k$ 成立. 第一部分证毕.

2) 证明 k 充分大时，$(\overline{x}^k,\overline{\lambda}^k)$ 可被 $\mathcal{F}_k$ 接受.

设 $\eta^k=\min\limits_{(p^l,f^l)\in\mathcal{F}_k}p^l>0$, 由于 $\overline{d}^k$ 是 $QP_s(x^k,\Delta^k)$ 的解，故 $g_i(x^k)+\nabla g_i(x^k)^{\mathrm{T}}\widehat{d}^k\le 0$, 于是

$$\begin{aligned}
\overline{p}^k&=p(x^k+\widehat{d}^k,\overline{\lambda}^k)\\
&=\max_{i\in I}\left|\sqrt{g_i^2(x^k+\widehat{d}^k)+(\overline{\lambda}_i^k)^2}+g_i(x^k+\widehat{d}^k)-\overline{\lambda}_i^k\right|\\
&=\max_{i\in I}\left|\sqrt{[g_i(x^k)+\nabla g_i(x^k)^{\mathrm{T}}\widehat{d}^k+\frac{1}{2}(\widehat{d}^k)^{\mathrm{T}}\nabla^2 g_i(y_i)\widehat{d}^k]^2+(\overline{\lambda}_i^k)^2}\right.\\
&\qquad\left.+g_i(x^k)+\nabla g_i(x^k)^{\mathrm{T}}\widehat{d}^k+\frac{1}{2}(\widehat{d}^k)^{\mathrm{T}}\nabla^2 g_i(y_i)\widehat{d}^k-\overline{\lambda}_i^k\right|\\
&\le\frac{1}{2}|(\widehat{d}^k)^{\mathrm{T}}\nabla^2 g_i(y_i)\widehat{d}^k|+\overline{\lambda}_i^k+(\widehat{d}^k)^{\mathrm{T}}\nabla^2 g_i(y_i)\widehat{d}^k-\overline{\lambda}_i^k\\
&\le|(\widehat{d}^k)^{\mathrm{T}}\nabla^2 g_i(y_i)\widehat{d}^k|.
\end{aligned}\tag{20}$$

此处 y 表示 x^k 到 $x^k+\widehat{d}^k$ 的连线上的一点. 由 $|\nabla^2 g_i(y_i)|\le M,\|\widehat{d}^k\||\le\Delta^k$ 于是 $\overline{p}^k\le M(\Delta^k)^2$. 即当 $\Delta^k<\sqrt{\frac{\beta\eta^k}{M}}$ 时，$(\overline{x}^k,\overline{\lambda}^k)$ 被 F 接受. 若 Δ^k 不满足此要求时，由算法 A 知，Δ^k 会不断减小，直至此要求满足.

综合 1) 和 2) 知，当 k 充分大时，必有 $(\overline{x}^k,\overline{\lambda}^k)$ 被 $\mathcal{F}_k\cup(x^k,\mu^k)$ 接受．引理证毕．

引理 4.2 算法 A 中循环 2 有限终止．

证 假设结论不真，由算法 A 知，存在子序列 $\{k\}\subset K$(k 为无穷序列), 使得 $\Delta^k\to 0(k\in K,k\to\infty)$ 且成立

$$\frac{\Delta\overline{f}^k}{\Delta\overline{q}^k}<\tau,\quad k=1,2,\cdots. \tag{21}$$

由 Taylor 展开式得

$$|\Delta\overline{f}^k-\Delta\overline{q}^k|=O(\|\widehat{d}^k\|^2)=O((\Delta^k)^2),$$

由 [11] 中引理 13.1.3 可得 $\Delta\overline{q}^k\geq\frac{1}{2}\|\nabla f(x^k)\|\Delta^k$, 于是

$$\Big|\frac{\Delta\overline{f}^k}{\Delta\overline{q}^k}-1\Big|=\frac{|\Delta\overline{f}^k-\Delta\overline{q}^k|}{|\Delta\overline{q}^k|}\leq\frac{2O((\Delta^k)^2)}{\Delta^k\|\nabla\overline{f}(x^k)\|}. \tag{22}$$

因 $\Delta^k\to 0,f\in C^2,x^k\in S$, 故可得 $\big|\frac{\Delta\overline{f}^k}{\Delta\overline{q}^k}-1\big|\to 0$, 这说明当 k 充分大时，必有 $\frac{\Delta\overline{f}^k}{\Delta\overline{q}^k}\geq\tau$, 这与 (21) 式矛盾．

将引理 4.2 中的 $\Delta\overline{f}^k,\Delta\overline{p}^k$ 分别用 $\Delta f^k,\Delta p^k$ 替换，即可得引理 4.3.

引理 4.3 算法 A 中循环 3 有限终止．

由引理 4.1– 引理 4.3 可看出算法 A 是可执行的，且在迭代过程中或者 $x^{k+1}=x^k+d^k$ 或者 $x^{k+1}=x^k+d^k+\overline{d}^k$ 成立．

下证算法 A 的全局收敛性.

为了更好地说明算法的收敛性，定义集合 Z 如下

$$Z=\{k|(x^k,\mu^k)\text{加入滤子集合}\}.$$

定理 4.4 若假设 A1–A4 成立，则由算法 A 产生的序列有如下情况

1) 迭代到问题 (P) 的 KKT 点;

2) 至少存在一个聚点是问题 (P) 的 KKT 点．

证 仅需对第二种情况给出证明．由引理 4.2, 引理 4.3 知，成功迭代的次数为无限次，又由 A2,$\{x^k\}\in S$, 故必存在聚点 x^*, 不妨设 $\lim\limits_{k\in K,k\to\infty}x^k=x^*$, 其中 K 为无限集．不失一般性，假设这样的无限循环出现在解 $QP_s(x^k,\Delta^k)$ 的分支上，由算法 A 可知，会出现如下两种情形

情形 A 有无穷多点加入滤子，即 $|Z|=\infty$.

i) 由加入滤子的条件可知 $\overline{p}^k>0$, 据引理 3.2, $\overline{p}^k\to 0(k\in K,k\to\infty)$. 即得 x^* 为可行点且满足 $\mu_i(x^*)\geq 0,g_i(x^*)\leq 0,\mu_ig_i(x^*)=0,\quad i\in I$, 其中 μ^* 为相应的拉格朗日乘子．

ii) 若 x^* 不是 KKT 点，则存在无限指标集 $K_1=\{k|\nabla f(x^k)^{\rm T}\widehat{d}^k>-\frac{1}{2}(\widehat{d}^k)^{\rm T}H^k\widehat{d}^k\}$, 若 $\exists K_2\subset K_1$ 使 $\lim\limits_{k\in K_2,k\to\infty}\|\widehat{d}^k\|=0$, 则可得 x^* 为 KKT 点，矛盾．故设存在 $\varepsilon>0$, 使得 $\forall k\in K_1$, 有 $\|\widehat{d}^k\|>\varepsilon$.

由子问题 $QP_s(x^k,\Delta^k)$ 的 KKT 条件知

$$\nabla f(x^k)^{\rm T}\widehat{d}^k=-(\widehat{d}^k)^{\rm T}\nabla g(x^k)^{\rm T}\overline{\lambda}^k-(\widehat{d}^k)^{\rm T}H^k\widehat{d}^k=(\overline{\lambda}^k)^{\rm T}g(x^k)-(\widehat{d}^k)^{\rm T}H^k\widehat{d}^k. \tag{23}$$

当 $g_i(x^k) \le 0$ 时，有

$$\nabla f(x^k)^{\mathrm{T}}\widehat{d}^k \le -(\widehat{d}^k)^{\mathrm{T}}H^k\widehat{d}^k \le -\frac{1}{2}(\widehat{d}^k)^{\mathrm{T}}H^k\widehat{d}^k. \tag{24}$$

当 $g_i(x^k) > 0$ 时，由 (16) 式有

$$\nabla f(x^k)^{\mathrm{T}}\widehat{d}^k \le (\overline{\lambda}^k)^{\mathrm{T}}\overline{p}^k - (\widehat{d}^k)^{\mathrm{T}}H^k\widehat{d}^k. \tag{25}$$

由 $\|\overline{\lambda}^k\| \le M, \overline{p}^k \to 0$ 知 $\exists k_0$, 当 $k \ge k_0$ 时，$\overline{p}^k \le \frac{1}{2M}(\widehat{d}^k)^{\mathrm{T}}H^k\widehat{d}^k$, 于是

$$\nabla f(x^k)^{\mathrm{T}}\widehat{d}^k \le -\frac{1}{2}(\widehat{d}^k)^{\mathrm{T}}H^k\widehat{d}^k. \tag{26}$$

而 K_1 为无穷集与当 $k > k_0$ 时 (24) 和 (26) 成立相矛盾，故为 KKT 点.

情形 B　有有限多点加入滤子，即 $|Z| < \infty$.

i) 此时由假设及算法 A 知，$f(x)$ 是单减且下有界的, 故由引理 3.1 知 $\overline{p}^k \to 0(k \in K, k \to \infty, K$为无限集), 这说明 x^* 为可行点，且满足 $\mu_i^* \ge 0, g_i(x^*) \le 0, \mu_i g_i(x^*) = 0,\ i \in I$, 其中 μ^* 为相应的拉格朗日乘子.

ii) 由算法 A 知, 存在 $k_0, \forall k \ge k_0$, $\nabla f(x^k)^{\mathrm{T}}\widehat{d}^k \le -\frac{1}{2}(\widehat{d}^k)^{\mathrm{T}}H^k\widehat{d}^k$ 且 $\Delta\overline{f}^k \ge \tau\Delta\overline{q}^k$, 又 $(\overline{x}^k, \overline{\lambda}^k)$ 被 $\mathcal{F}_k \cup (x^k, \mu^k)$ 所接受，故有 $f^k - f^{k+1} \ge \gamma p^{k+1} = \gamma O(\|\widehat{d}^k\|^2)$, 两边求和

$$\sum_{k=k_0+1}^{\infty}(f^k - f^{k+1}) \ge \sum_{k=k_0+1}^{\infty}\gamma O(\|\widehat{d}^k\|^2).$$

而左边是有限数，故有 $\|\widehat{d}^k\| \to 0(k \to \infty)$, 从而 x^* 为 KKT 点.

当无限多的成功迭代出现在子问题 $QP(x^k, \Delta^k)$ 的分支上时，在如上的证明中用 d^k 代替 $\widehat{d}^k$ 即可得相同结论.

参　考　文　献

[1] Fletcher R and Leyffer S. Nonlinear programming without a penalty function. *Math. Program.*, 2002, **91**: 239–269.

[2] Fletcher R, Leyffer S and Toint P L. On the global convergence of a filter-SQP algorithm. *SIAM J. Optim.*, 2002, **13**: 44–59.

[3] Wachter A and Biegler L T. Global and local convergence of line search filter method for nonlinear programming. Department of Chemical Engineering Carnegie Mellon University, Pittsburgh, 2002.

[4] Zhang Juliang and Wang Yong. A new trust region method for nonlinear equations. *Math. Meth. Oper. Res.*, 2003, **58**: 283–298.

[5] Ulbrich S. On the superlinear local convergence of a filter-SQP method. *Math. Program.*, 2002, **100**: 217–245.

[6] Fletcher R, Gould N I M, Leyffer S, Toint P L and Wachter A. Global convergence of a trust region SQP-filter algorithm for general nonlinear programming. *SIAM J. Optim.*, 2002, **13**: 635–660.

[7] Fletcher R, Leyffer S and Toint P L. On the global convergence of an SLP-filter algorithm. Department of Mathematics, University of Namur, Namur, Belgium, 1998.

[8] Nie P Y. Sequential penalty quadratic programming filter methods for nonlinear programming. *Nonlinear Analysis: Real World Applications*, 2007, **8**: 118–129.

[9] Nie P Y. A compsite-step type filter mehtod for equality constrained problems. *J. Comput. Math.*, 2003, **21**: 613–624.

[10] Peng Y H and Liu Z H. A derivative-free filter algorithm for nonlinearity problem. *Appl. Math. Comput.*, 2006, **182**: 846–853.

[11] 袁亚湘，孙文瑜．最优化理论与方法．北京：科学出版社， 1997.

TRUST-REGION FILTER METHOD WITH NCP FUNCTION

SU Ke

(*College of Mathematics and Computation, Hebei University, Baoding* 071002;

Department of Mathematics, Tongji University, Shanghai 200092)

Abstract Filter method was initially proposed by Fletcher and Leyffer in 2002. If the objective function value or the constrained violation is reduced, this trial point is accepted, which is the basic idea of the filter method. In this paper, the Fischer-Burmeister NCP function value is used to modify the violation function value in the filter. It is shown that the new filter method has the global convergence property.

Key words Filter, trust-region, nonlinear complementarity, convergence.